

音声認識技術を活用した聴覚障害学生支援：教室内及び遠隔地における通訳システムの構築

著者	松崎 丈，藤島 省太，田幡 憲一，安藤 明伸，前原 明日香，及川 麻衣子
雑誌名	宮城教育大学情報処理センター年報
号	16
ページ	A1-A6
発行年	2009-03-31
URL	http://id.nii.ac.jp/1138/00000369/



音声認識技術を活用した聴覚障害学生支援 ー教室内及び遠隔地における通訳システムの構築ー

松崎 丈¹，藤島 省太¹，田幡 憲一²，安藤 明伸³，前原明日香⁴，及川麻衣子⁴

¹宮城教育大学 特別支援教育講座

²宮城教育大学 理科教育講座

³宮城教育大学 技術教育講座

⁴宮城教育大学 教務補佐員（障害学生支援コーディネーター）

本研究は、聴覚障害学生が授業で教員や受講生の音声情報を獲得することを保障するために、音声認識技術を活用した聴覚障害学生支援を試みたものである。教室内で通訳する場合と遠隔地から通訳する場合の2つにわけて、音声認識技術を活用した通訳システムを構築し、聴覚障害学生支援を行った。その結果、利用者である聴覚障害学生から、教室内と遠隔地の両方とも、手書きノートテイクと比べて情報量が多く、授業に参加できているという実感があるとの評価がえられた。その一方で、システム構築の安定性・効率化や、教員の話し方や学生の活動がある授業への対応方法が課題になった。

キーワード：聴覚障害学生支援，音声認識技術，教室内，遠隔地，音声認識通訳システム

1. はじめに

宮城教育大学は、大学のユニバーサルデザインの実現を目指して、障害学生支援プロジェクトを組織し、学生との相互協力によって障害学生支援のシステムを構築するとともに、障害学生への学習・生活支援、環境整備等に積極的に取り組んでいる[1]。

2. 聴覚障害学生支援の現状と課題

障害学生支援は、障害学生の自主的な申請によって始めている。障害学生の大部分を占める聴覚障害学生7名のうち6名は、授業における文字通訳の支援を申請している。従来行われている文字通訳は、教員の音声情報を約20%の要約率で筆記する手書きノートテイクであり、聴覚障害学生が健聴学生と対等に教育を受ける権利を保障するという観点から見れば充分とはいえない状況であった。

そこで、教員の音声情報をより多く文字化していくために、平成19年度に採択された新たな社会的ニーズに対応した学生支援プログラムの事業「障害学生と共に学べる総合的学生支援（事業責任者：藤島省太）」で新たな文字通訳の技術を導入することとした。すなわち、パソコンノートテイクと音声認識技術を活用した通訳（以下「音声認識通訳」）である。特に音声認識通訳は、近年の聴覚障害者支援における最新技術として注目されており、今後、支援現場における導入と実践に関する基礎資料の蓄積が課題となっている。

本研究では、聴覚障害学生支援において音声認識技術を活用してどのような通訳システムを構築したのか、授業現場にどのように導入し、そこでえられた成果や課題は何であったかを報告する。

3. 音声認識技術について

音声認識とは、音声情報を自動的に認識して字

幕化する技術である。パソコンノートテイクでは、話者の音声情報の40～80%程度字幕化されるが、音声認識では要約からほぼ全文までの字幕化が可能である[2]。つまり、音声認識技術を活用すれば、手書きノートテイクやパソコンノートテイクでは要約・省略される”文字で表記できる狭義の言語情報”をほぼ全て字幕化でき、かつ話し手の判断や心的態度も把握できるような”話しことば固有の構文情報（省略・繰り返し・言い直し・挿入・倒置等）”も字幕化されることが可能となる。これにより、聴覚障害学生に「臨場感あふれる[3]」形で伝えられることが期待される。

ただし、現行の音声認識技術では、音声情報を100%正しく認識して字幕化することが不可能である。また、誤認識される字幕には、元の音声内容を推測できないものがあることに注意しておく必要がある。例えば、字幕「この坂俳句ですか」は、元の音声情報「この魚はいくらですか」を誤って認識されたものである。このような誤認識は、音声情報の内容がわからない聴覚障害学生にとって読みにくくなり、情報保障の質を低下してしまうことにつながりかねない。

4. 音声認識通訳システムの概要

上記の音声認識技術の技術的な限界を考慮して図1のように音声認識通訳システムを構築した。

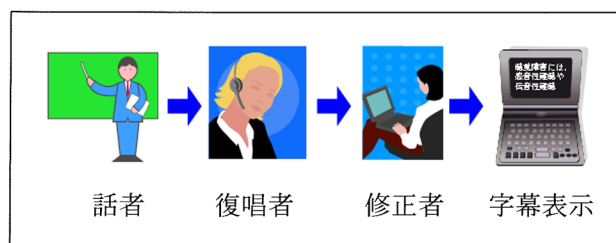


図1 音声認識通訳システムの概要

図1の話者が音声情報を発信すると、復唱トレーニングを受けた復唱者が、話者の音声情報を音声認識ソフトに認識されやすい音声情報に変換して復唱用パソコンに音声入力する。復唱者は、一定の発話速度と明瞭度を維持しながら、話者の音声情報のほぼ全文を復唱する。したがっ

て、教員の側にも、比較的ゆっくりした発話速度で説明するなどの配慮が必要になると考えられる[4]。教員の発話速度が速くなると、復唱者はほぼ全文での復唱は難しいため、その結果、パソコンノートテイクと同じように内容を要約して復唱せざるをえなくなる。この点で今後、復唱者が復唱しやすい教員の発話方法に関するFD研究をする必要性はあると思われる。復唱作業に用いる音声認識ソフトウェアについて、最初、市販の音声認識ソフトウェア「ドラゴンスピーチ 2005 Professional」を活用していたが、最近では「AmiVoice Es 2008」を用いている。理由は、前者は音声入力をする者の音声データを事前登録する必要があるのに対し、後者は音声認識技術のさらなる向上で事前登録が不要になり、かつ専門用語をほぼ正確に認識することが可能だからである。

図1で復唱者の音声入力によって認識された字幕には、誤認識が約10%以上は生じている。そこで、復唱者用パソコンから修正者用パソコンにネットワーク経由で自動的に送った字幕を修正者が修正して、字幕の認識精度を高める必要がある。

まず、復唱者のパソコンで産出した字幕を修正者のパソコンに自動的に送る必要があるが、そのために「SR - MODULE」というソフトウェアを復唱者用パソコンにインストールして実施したが、最近では「SR - LAN2dash」に変更した。このソフトウェアは、復唱者用、修正者用、表示用の各パソコンにインストールして、パソコン同士の字幕送受信に用いるもので、ネットワークの安定性が高い。「SR - MODULE」と「SR - LAN2dash」のいずれも、筑波技術大学で音声認識通訳を研究しているグループが開発し、今回の導入のために無償提供されたものである。

次に、修正作業に使うソフトウェアについて、最初は「IPtalk9t (<http://iptalk.hp.infoseek.co.jp/>)」を用いていたが、最近変更した「SR - LAN2dash」に修正機能が追加されたため、後者を用いている。

さらに、その修正作業の段階で、教員の音声も同時に聞いて修正できるように、音声を数秒ほど遅延させて修正者のヘッドホンに送信する音声遅延システムを導入する。この音声遅延に用いている機器は、(株)B.U.G.の音声遅延装置Videoboxである(現在は生産中止)。なお、授業は90分間であるため、復唱者や修正者が長時間で一定の認識精度を維持して音声認識通訳を行なうために、復唱者は基本的に2名体制とし、10分交代で行い、修正者は1名体制で行なった。

こうして話者の音声情報を約95%以上の認識精度で字幕化した結果は、ネットワーク経由で聴覚障害学生及び教員がそれぞれ見る表示用パソコンに表示される。話者の音声は復唱・修正作業を通して字幕がどのように作られるのかについて、表1に示す。

表1 音声認識通訳の過程における字幕作成

1. 話者の音声

実際に測定すると皆さんの場合も右の耳と左の耳が全く同じという人は少ないと思います。って、その0の方が、えー、本当は、えー、音があるんだけど、皆さんがきくと初めて聞こえたところ感じる点になるということです。

2. 復唱作業によって認識された字幕結果

実際に測定をすると、みなさんの場合も、新富右と左が全く同じという人は少ないと思います。ゼロというのは、本当は、夫があるのだけれども、皆さんが聞くと、初めてことが聞こえたというところです。

3. 修正作業によって呈示された字幕結果

実際に測定をすると、みなさんの場合も、右と左が全く同じという人は少ないと思います。ゼロというのは、本当は、音があるのだけれども、皆さんが聞くと、初めて音が聞こえたというところです。

このような音声認識通訳システムを構築するために必要とされる機材は表2の通りである。なお、復唱者及び修正者が音声認識通訳のために行う支

援技術の詳細については、文献[4]を参照されたい。

表2 音声認識通訳システムの構築に使用した機材

機 材	数 量
復唱者用パソコン	2
復唱者用マイク	2
復唱者用サウンドブラスター	2
修正者用パソコン	2
修正者用密閉型ヘッドホン	2
ヘッドホンアンプ	1
音声遅延装置	1
表示用パソコン	2
スイッチング HUB	1

5. 授業における音声認識通訳システムの導入

平成20年度において、宮城教育大学で、表3のように3つの授業に音声認識通訳を導入した。

表3 音声認識通訳を導入した授業

	授 業	開講	教室	担当教員
①	聴覚・言語障害の心理・生理・病理	前期	730	藤島省太
②	聴覚・言語障害への教育支援B	後期	720	藤島省太
③	理科教材研究法c	後期	理科共用実験室	田幡憲一

授業に音声認識通訳を実施するにあたって、授業①と②では、復唱者及び修正者を教員・聴覚障害学生と同じ教室(以下「教室内」)に配置して音声認識通訳を行った。一方で、授業③では、授業担当教員の提案により、復唱者及び修正者を教室とは別室(以下「遠隔地」)に配置し、遠隔地からLANネットワークを介して音声認識通訳を行うこととした。

以下、教室内及び遠隔地において音声認識通訳システムをどのように構築したのか、システム構築及び音声認識通訳の実施について実際にどのような課題が挙げられたかを報告する。

5.1 教室における音声認識通訳の実施

授業に音声認識通訳システムを導入する際、教員及び支援者にとって初めての取り組みになるため、授業①では、お互いの行動を観察しながら通訳作業を調整できるように、支援者も教室内で音声認識通訳を行うことにした。教室内で構築する音声認識通訳システムは、前節で述べたシステムと同じである。ただし、復唱作業中に復唱者以外の音声が入力されたり復唱者の音声を受講生に聞かれることを防ぐために、通常の指向性マイクではなく、復唱者の鼻と口を覆うマスク型マイク（TALK TECHNOLOGIES SM300）を活用した（図2）。



図2 マスク型マイク

このマイクは、指向性マイクより認識精度は若干低下するものの[5]、教室の様子を見ながら通訳できるというメリットがある。ただし、復唱者が自分自身の音声を聞くことができなくなるため、音声認識ソフトウェアの認識結果を見ながら復唱時の発話を調整する必要がある。上記のマスク型マイクを前項で述べた音声認識通訳システムに組み合わせたものを教室内でセッティングして音声認識通訳を実施した。

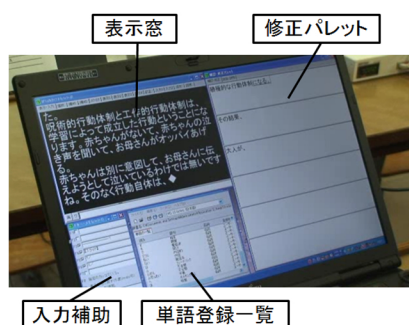


図3 修正者用パソコンの画面

なお、授業①において、当初は復唱者及び修正者が音声認識通訳の技術を習得中であったため、話者の音声情報が字幕化されて表示用パソコンに表示されるまでの時間で約10秒以上要する 때가あった。聴覚障害学生はできる限り早く情報を獲得したいというニーズを持っていたため、表示用パソコンだけでなく、修正者用パソコンの修正パレットも見たいとの要望が出された。修正用パソコンの画面には、図3のように修正パレットがあり、復唱者の音声入力によって字幕化された結果が修正パレットの空欄に一句一節単位で順次表示される。修正者は順次表示された字幕結果を見て修正する。なお、図3は、SR-MODULEとIPTalk9tを使っている時である。聴覚障害学生が、前列にある修正者用パソコンと聴覚障害学生が用いる表示用パソコンの両方を見れるように、各パソコンを図4のように配置した（図内の矢印は聴覚障害学生の視線方向であることを示す）。また、表示用パソコンで字幕フォントのサイズを小さくするとともに、1行あたりの文字数、行数を通常の設定より増やして本を読んでいるようなイメージで表示した[5]。これは、当初の音声認識通訳の運用状況に応じて利用者のニーズを導入することで、音声認識通訳システムを柔軟的に構築した例である。



図4 各パソコンの配置レイアウト

このように教室内で音声認識通訳を実施した結果、音声認識通訳システム構築及び音声認識通訳について次のような成果と課題が確認された。

(1) 授業担当教員は、少しゆっくりした発話速度で話し、時々字幕作成の状況を表示用パソコン

で確認して大幅に遅れている場合は待つ行動をとることがあった。これにより復唱者は一定の発話速度で明瞭に復唱作業を行うことが可能となり、聴覚障害学生への情報保障の質が保障された。

(2) 授業①や②を受講した聴覚障害学生 A から、手書きノートテイクと比べて、授業からえられる情報がとて多くなり、満足しているとの評価が出された。

(3) 同授業を受講した一般学生から、音声認識通訳システムに驚嘆し、聴覚障害学生への支援の必要性が認識できたとの反応がえられた。情報保障の必要性を共有する契機になったと思われる。

(4) 課題として、教室内で音声認識通訳を行うためには、教室内に4名ほどの支援者や支援機材を配置できるスペースが確保されること、かつ支援機材の運搬及びシステムのセッティングのため授業前に約20分間の準備時間が保証されることの2つの条件が必要であることが挙げられた。

5.2 遠隔地における音声認識通訳の実施

平成20年度後期から、教室内での音声認識通訳に加えて(授業②)、遠隔地における音声認識通訳も実施した(授業③)。後者は別室で音声認識通訳システムを構築して、教室内の教員の音声情報を別室へネットワークで送って音声認識通訳を行い、産出された字幕結果をネットワークで折り返し送信して、教室内の表示用パソコンに出すという方法である。

そうした遠隔地における音声認識通訳を実施するために、本学の学生支援GP事業の一部として、宮城教育大学情報処理センターの協力のもと、LANネットワークがない教室にネットワーク整備の工事を行なった。その結果、3号館、4号館、5号館、6号館、8号館、9号館、音楽棟、理化学棟、萩朋会館の各教室にLANネットワークが整備された。

次に、授業③において、理科共用実験室内の教員の音声や映像をLANネットワークで送るために、鳴原・堀越・田幡(2008)[6]のWebカメラを活用したネットワークシステムを参考にした

り技術教育講座の安藤の技術的なサポートを受けて次のようなシステムを構築した。理科共用実験室内で、授業担当教員は、ヘッドホンマイク(Panasonic X004359 U01)とワイヤレスマイクロホン(Panasonic WX4300B)を装着する。理科共用実験室内に、その音声情報を受信するポータブルアンプ(Panasonic WX-282C)とマルチターミナルモジュール(Canon VB-EX50)をおく。この受信された教員の音声と一緒に教員の行動も映像配信できるネットワークカメラ(Canon VB-C50i)につなげて、音声認識通訳システムが設置されている遠隔地にLANネットワークで送信する。

遠隔地における音声認識通訳は、宮城教育大学3号館3階の特別支援教育総合研究センター客員研究員室で行うこととし、授業が始まるまでに音声認識通訳システムを立ち上げ、復唱者2名及び修正者1名の体制で支援を行った。なお、理科共用実験室内でシステムの不具合や支援の技術的限界などが生じた場合を想定して、教室内にサブテイカーを配置して補助的支援を行った。

このように遠隔地で音声認識通訳を実施した結果、音声認識通訳システム構築及び音声認識通訳について次のような成果と課題が確認された。

(1) 授業担当教員から、教員が一方的に話し始めた場合に話す速度が速くなるため復唱者が追いつけなくなることがあったとの報告が出された。教員及び復唱者がお互いに負担を軽減して進められるように、今後、教員や支援者はどのような配慮をしたらよいのかFD研究を進めていく必要があると思われる。

(2) 授業③を受講した聴覚障害学生 B から、次のような評価がえられた。授業担当教員が音声認識通訳を受け入れたおかげで、文字量はノートテイクより多く、迅速にえられて授業を受けている実感があるとのことであった。今後は、教員の話し方や実験活動で生じる学生の発言に対してどのように対処するかを皆で検討する必要があると考えているとのことである。

(3) 遠隔地による通訳システムであるため、支

援者が通訳する上で次のような検討事項が確認された。1つは、ネットワークカメラの映像からえられる映像情報が限られているため、教員が指示語の内容を説明したり、板書の説明に図に番号を振って図の番号を言いながら説明するなどの工夫が必要になること。もう1つは、学生の活動が伴う実験形式の授業があるため、復唱者が学生の様子を観察できるようにもう1台のネットワークカメラを通常より高い位置に設置する必要があるか、もしくは教員が学生の発言内容や様子を話すなどの工夫が必要であること。なお、ガンマイクを使って学生の音声情報を拾うことも試みたが、実際は十分に拾えなかった。

(4) SR - MODULE と IPtalk9t を使っている時は、作成した字幕を理科共用実験室の表示用パソコンにうまく表示できない問題が度々生じていたが、SR - LAN2dash に変更してからは改善された。

(5) 遠隔地による通訳では、必要な機材が増加するため、教室内で行う場合と比べて機材を運ぶ手間はなくなる反面、いかに効率良くセッティングするかが課題となった。遠隔地と教室内であらかじめ簡単にセッティングできるようなシステムの開発が必要である。

6. おわりに

本研究では、教室内及び遠隔地における音声認識通訳システムの構築及び実践について報告した。それぞれの支援を受けた聴覚障害学生 A, B からは、情報保障が質的に向上したとの肯定的な評価がえられた。一方で、教室内及び遠隔地において音声認識通訳システムをより効率的に構築すべき課題があることが確認された。また、音声認識通訳の質的保障には、教員の話し方や学生とのやりとりが含まれる授業形態が大きく関与していることが明らかになったため、今後、FD 研究と関連付けて検討していく必要がある。今後も、聴覚障害学生だけでなく、支援者、授業担当教員、受講生にとっても使いやすい音声認識通訳システムの構築を目指して取り組んでいきたいと考

える。

7. 謝辞

本研究は、文部科学省の平成 19 年度「新たな社会的ニーズに対応した学生支援プログラム」に選定された宮城教育大学の「障害学生も共に学べる総合的学生支援－障害学生との共生により人間性豊かな社会人を育成するための入学から就職までの総合的学生支援システム構築－（事業推進責任者 藤島省太）」の一部として実施したものである。

参考文献

- [1] 藤島省太, 松崎丈: 教員養成大学の特色を生かした障害学生支援体制の構築－宮城教育大学の障害学生支援プロジェクト, 大学と学生, 46, pp.31-38, (2007).
- [2] 三好茂樹: PEPNet-Japan Tipsheet 音声認識技術を用いた情報保障, 日本聴覚障害学生高等教育支援ネットワーク, (2008).
- [3] 中野聡子, 牧原功, 金澤貴之, 菊池真里, 黒木速人, 井野秀一, 伊福部達, 福島智: 音声認識技術を利用した字幕呈示システムの現状と課題－音声言語と文字言語の性質の違いに焦点をあてて, 群馬大学教育実践研究, 23, pp.251-259, (2006).
- [4] 松崎丈, 藤島省太: 聴覚障害学生支援における音声認識を活用した通訳システムの構築－利用者の視点に観点に基づいた字幕提示の検討－, 宮城教育大学紀要第 43 巻, (印刷中).
- [5] 三好茂樹, 黒木速人, 河野純大, 白澤麻弓, 石原保志, 小林正幸: 音声認識技術を利用した字幕作成担当者のための支援技術とそのシステム開発, 筑波技術大学テクノレポート, 14, pp.145-152, (2007).
- [6] 鳴原卓, 堀越詩野, 田幡憲一: WEB カメラを用いた学生支援システムの構成－附属小学校と学部を結ぶ－, 宮城教育大学情報処理センター年報, 第 15 号, F-1－F-6, (2008) .